# Recent Accomplishments in Coupling High Resolution Earth System Models Using Advanced Computational Technologies

Joseph L Eastman [1], Christa Peters-Lidard [2], Wei-Kuo Tao [3], Sujay Kumar [1]. Xiping Zeng [1], Steve Lang [4]

1. University of Maryland Baltimore County, Goddard Earth Sciences and Technology Center, Baltimore, Maryland, USA
2. Hydrological Sciences Branch NASA-GSFC Code 614.3, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA
3. Laboratory for Atmospheres, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA
4. Science Systems and Applications, Inc. Lanham, Maryland, USA

*Abstract-* **The NASA/GSFC Land Information System (LIS) has been coupled to the Weather Research and Forecasting (WRF) and Goddard Cumulus Ensemble (GCE) models using advanced parallel and Earth System Modeling Framework (ESMF) techniques. LIS is a high-performance Land Data Assimilation System (LDAS), developed under funding from ESTO-CT's Round-3 CAN, whose TRL3 technologies have been advanced to TRL4, as well as the coupled LIS and WRF (LISWRF) system. Recently LIS has been updated to include multiple nests and projections. The uncoupled LIS is first integrated over a long period of time using a combination of forcing and soil datasets with different to provide the initial land surface conditions for LISWRF. LISWRF is also executed with the default land surface state from WRF Standard Initialization (WRFSI). LISWRF is then integrated with these initial conditions (ICs) for two separate case studies. The results show significant improvement in simulating convective events for all LIS derived ICs versus WRFSI, while higher resolution forcing and soil data used in LIS integrations exhibits mixed results. In a separate study LIS derived fluxes were used to drive the GCE model and compared with those from the Atmospheric Radiation Measurement (ARM) data. Here was found that the LIS derived fluxes improved the ability of GCE to simulate diurnal cloud variation in the lower troposphere.**

## I. INTRODUCTION

The effect of various land surface data assimilations as input to a meteorological model's initial surface conditions and the subsequent impact on convective modeling is investigated for two case periods with significant mesoscale convection. The periods, June 12-13 and May 24-25, 2002 were intensively observed during the International H2O Project (IHOP) field experiment [16]. Several past studies have shown the effects of surface heterogeneity and soil moisture gradients on the development of convection (e.g. [3], [9], [12]). In [3] the authors demonstrated the influence of the effects of vegetation and soil processes on convection and suggest that a detailed representation of these processes should be included in forecast models. A recent discussion [8] suggests that the role of heterogeneity is of such importance that it needs to be included in future climate models.

In recent years, considerable amount of effort has been devoted to developing realistic representation of land surface boundary in coupled simulations. NASA's Goddard Space Flight Center has developed a Land Information System (LIS [5], [4], [14]) capable of simulating global land surface conditions at spatial resolutions down to 1km. LIS is a high performance Land Data Assimilation System (LDAS [10]) that consists of several land surface models run offline using observationally-based precipitation, radiation, and meteorological inputs, and surface parameters. LIS provides the capability to integrate land surface simulation, observation, and analysis methods to accurately determine land surface energy and water states. The fine spatial scales employed by LIS in generating the initial conditions for LISWRF will improve the ability the coupled system to simulate and eventually predict the initiation and evolution of precipitating systems.

## II. DESCRIPTION OF EXPERIMENTAL DESIGN

The objective of this study is to investigate the role of high resolution initial surface conditions, derived from different land surface integrations, on the development of convection for two synoptically different case periods and to assess the ability of LIS at simulating these conditions. The May 24-25 period represents a strongly forced convective case day, while June 12-13 was characterized by weaker forcing characterized by relatively weak winds and no upper level forcing. It would be expected that fine scale representation of the initial land surface conditions would have a stronger impact on the developing convection for the weakly forced case study. In this study we derive the initial conditions through integrations of the Land Information System (LIS, [4]) employing different resolutions of atmospheric forcing and soil representation.

For each case period studied there were a total of five integrations performed using LIS coupled to the Weather Research and Forecasting (WRF, [13]) model. This study used a total of 501x501 horizontal grid pints at 1km grid spacing for both LIS and LISWRF integrations and is displayed in Fig. 1. LISWRF also had 45 vertical sigma-z levels ranging from 10m at the surface to a model top of 19km for the June 12-13 study and 23km for the May 24-25

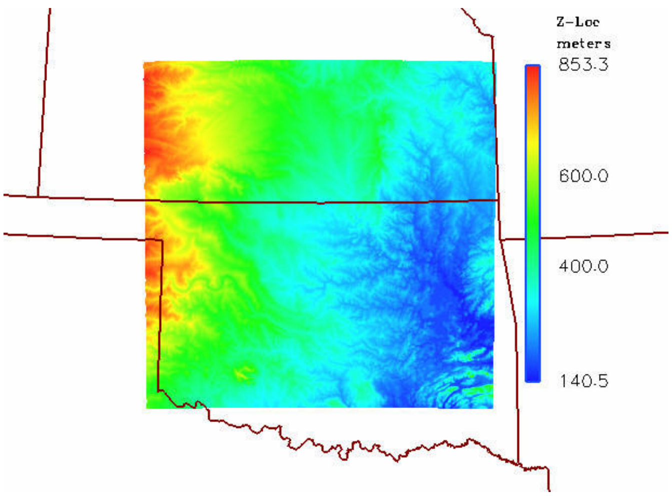


**Fig. 1: Domain used for all integrations. Colored fields indicate the topography.**

study. The atmospheric model was set up to with sophisticated microphysics that allows the model to resolve 5 different species. In addition, the shortwave and longwave parameterizations account for the effects of hydrometeors. The model timestep was set to 6 seconds for all integrations.

In order to provide the coupled modeling system with initial land surface conditions the LIS modeling was used to spin up these fields. For this study the NOAH [2] land surface model (LSM) was employed. This LSM simulates soil moisture (both liquid and frozen), soil temperature, skin temperature, snowpack depth, snowpack water equivalent (and hence snowpack density), canopy water content, and the energy flux and water flux terms of the surface energy balance and surface water balance. The LSM land-surface parameters were initialized with the University of Maryland 1KM data sets for vegetation and land-sea masks. Climatological data sets were ingested in order to initialize other vegetation parameters such as albedo and green vegetation fraction. Initial soil water and temperature profiles were also assigned according to climatology.

The LIS integrations employed combinations of meteorological forcing from the Global Data Assimilation System (GDAS, [1]) and the North American Land Data Assimilation System (NLDAS, [6]) in conjunction with soil data derived from the Food and Agriculture Organization (FAO, [10]) and the State Soil Geographic (STATSGO, [14]) database. The four spin up integrations are summarized in Table 1. The GDAS forcing has a 3 hour temporal resolution while NLDAS has 1 hour resolution.

| Simulation Symbol | Forcing | Forcing Resolution | Soils | Soil Resolution |
|---|---|---|---|---|
| F0 | GDAS | 2.5 degree, 3 hour | FAO | 5 minute |
| F1 | GDAS | 2.5 degree, 3 hour | STATSGO | 1km |
| F2 | NLDAS | 1/8$^{th}$ degree, 1 hour | FAO | 5 minute |
| F12 | NLDAS | 1/8$^{th}$ degree, 1 hour | STATSGO | 1km |

**Table 1: Integration symbol and combination of forcing and soils data used.**

In all integrations LIS was first integrated for the period 1987 through 1996 employing the National Center for Environmental Prediction reanalysis data (NCEP, [14]) for the atmospheric forcing after which the aforementioned atmospheric forcing was used. All integrations were evaluated from May 1 through June 26, 2002, the end of the IHOP field program in order to ascertain the skill of LIS at simulating the existing soil conditions and to assess the forcing data employed in LIS.

## III. Evaluation of LIS Integrations

We evaluated the LIS integrations against the station data of the Oklahoma mesonet (http://www.mesonet.ou.edu). The mesonet data is in 5 minute average intervals which were average up to 30 minute intervals. In total over 190000 observational records were compared to the LIS forcing data. Statistical quantities such as root mean squared error (RMSE), bias, average, variance and correlation were computed for the IHOP period. In terms of the prognostic variables evaluated there were minor differences in the soil temperature and moisture quantities with no indication of improvement for higher resolution forcing or soil data. In both cases the soils were hotter and drier than observed.

Surprisingly, there were no significant differences in the statistics for the various forcing variables. Correlations were all above 0.75 for the mixing ratio, temperature, wind speed, and pressure, with some as high as 0.98. Again, the atmospheric forcing was in general hotter and drier. The worst performing forcing variable was the precipitation, with a deficit bias of -0.29mm for GDAS, and -0.28mm for NLDAS. In both cases the correlations were well below 0.3. The rainfall deficit may not seem like a large amount, until one considers the number of events where the observations failed to agree with a precipitation event was roughly 30000 for NLDAS and 32000 for GDAS. Overall this will significantly impact the soil moisture state and implies the need for higher resolution precipitation data, which is currently being addressed with the incorporation of Stage IV bias corrected, radar derived precipitation data into LIS. This data has a 4km horizontal resolution and 1 hour temporal resolution.

## IV. CASE STUDIES

Since we are examining the impact of high resolution coupled models we first searched for a day that magnified the importance of the land surface conditions and resulting mesoscale circulations. In most field studies there are days referred to as "Golden Days". Clear skies and relatively weak synoptic forcing characterize these days. The enhanced solar radiation combined with weak synoptic forcing allows the mesoscale fluxes to develop maximum strength through differential heating. The differential heating can be a result of topography, soil moisture gradients, or land use patterns to name a few examples. The day that exhibited the desired properties during the IHOP campaign was June 12-13th, 2002. There was a significant moisture gradient in the western portion of the domain, a surface mesoscale low to the northwest, and outflow from a previous system that had exited the domain. Fig. 3 features interact to produce large convective system initiated near the west center of the domain at roughly 2130GMT on the 12th.

All integrations of LISWRF were started at 12GMT on June 12th, 2002 and simulated a 24 hour period. The only differences were the use of different LIS generated ICs and one integration with WRFSI ICs.

As in the LIS evaluation, the stations of the Oklahoma mesonet were used to evaluate the model simulated fields of mixing ratio, wind speed, temperature, and precipitation. Overall there were slight improvements as the forcing and soils resolution was increased. We also used the Stage IV, bias corrected radar derived precipitation product to compare gridded values for the integrations. In order to accomplish this LISWRF output were interpolated to the Stage IV grid locations. The final 24 hour spatial patterns of accumulated precipitation are shown in Fig. 4. The case of using default WRFSI for initial conditions clearly inferior to the integrations employing LIS generated ICs. This particular day displayed convection that was not organized in a line, but was considered discontinuous in its structure. This is represented in the observations and LISWRF integrations with LIS ICs by the patchy nature of the convection, where distinct areas of convection leave a NW to SW orientation in the streaks of precipitation.

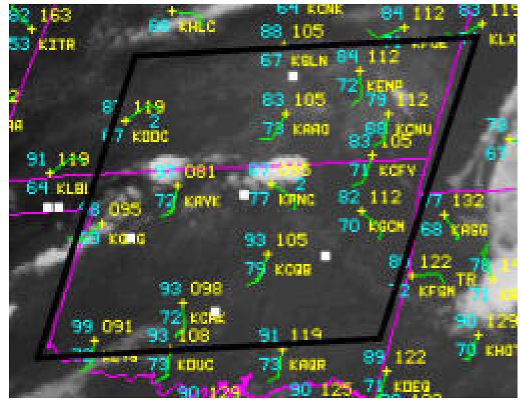

Fig. 3: GOES infrared imagery at 2130GMT on June 12th, 2002 overlaid with surface observations. Black lines indicate the approximate location of our modeling domain.

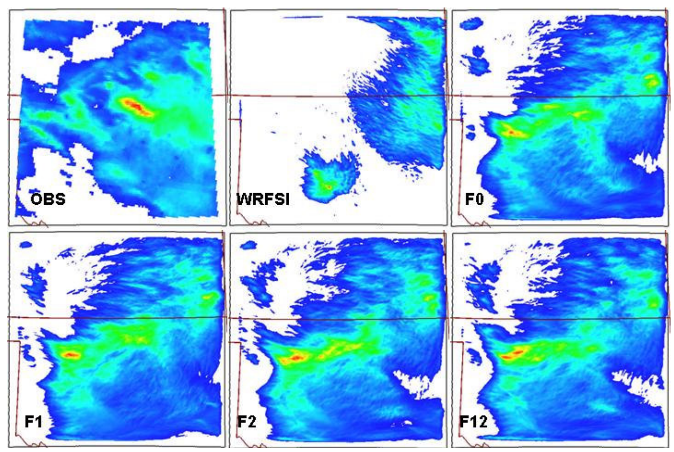


Fig. 4. 24 hour accumulated precipitation patterns for observations (top left), WRFSI (top center), F0 (top right), F1 (lower left), F2 (lower center), and F12 (lower right).

Fig. 5 displays the normalized domain integrated total precipitation values for the LISWRF integrations and observations. Overall, the temporal response of the LISWRF integration using the LIS generated ICs shows good temporal agreement with the observations in terms of the onset of the convective event. This in an improvement as most mesoscale models have a tendency to lag the actual timing of a convective event. The integrations do not build as quickly as the observations and the rainfall is continuing while the observations indicate a tailing off much sooner than modeled.

It appears that the LISWRF integrations continue to back-build and overestimate the observed precipitation. Overestimation is a common bias in high resolution convective modeling. The LSIWRF integration using WRFSI showed an underestimate by an order of magnitude and the

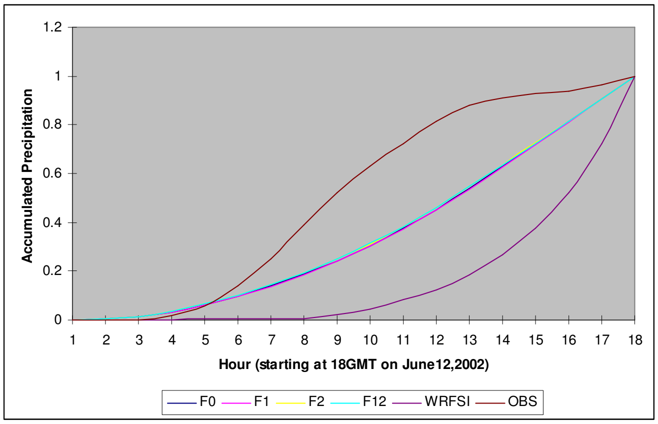


Fig. 5. Normalized accumulated precipitation totals starting at 18GMT on June12th, 2002 for LISWRF integrations and observations.

temporal agreement is clearly lagging. In this case WRFSI is using data with a spatial resolution of roughly 40km to initialize the soil moisture and temperature gradients. It is likely this reduction in the fine structure of these gradients leads to the subsequent reduction in these gradients and their influence on mesoscale fluxes.

Our second case day focused on May 24-26th, 2002. This day was characterized by strong synoptic forcing that featured a slow-moving cold front in conjunction with a upper level short wave trough that deepened over the integration period. This cold front extended from Texas-Oklahoma panhandle to northeastern Kansas and slowly propagated to the southeast. Again, there was a dryline in the southwest portion of the domain which interacts with the southeasterly flow of the moist air from the Gulf of Mexico. Since this is a strong synoptically forced system it would seem likely that the ICs would have a relatively minor impact when compared to the June 12th case.

Fig. 6 displays the spatial precipitation patterns after 48 hours of integration. Surprisingly, despite the strong forcing, the WRFSI integration fails to capture a majority of the observed precipitation, although the precipitation totals were closer to the LISWRF integrations. The LISWRF integrations exhibit some similarities to the observed precipitation, but failed to capture the large area of convection in the northeast portion of the domain. This was due to the atmospheric initialization, which did indicate as strong of a frontal boundary as was observed.

We again examined the temporal response of the model integrations compared to observations. The curves are shown in Fig. 7 for all integrations and the observed precipitation. The observations indicate a large accumulation of precipitation during the period 0GMT to 12GMT on the 24th. This is due to an existing system that had already developed at the start of the integration. Clearly the model could not

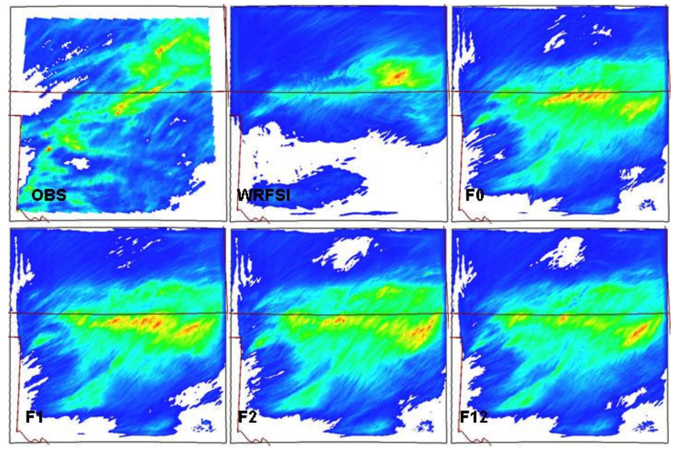


Fig. 6. 24 hour accumulated precipitation patterns for observations (top left), WRFSI (top center), F0 (top right), F1 (lower left), F2 (lower center), and F12 (lower right).

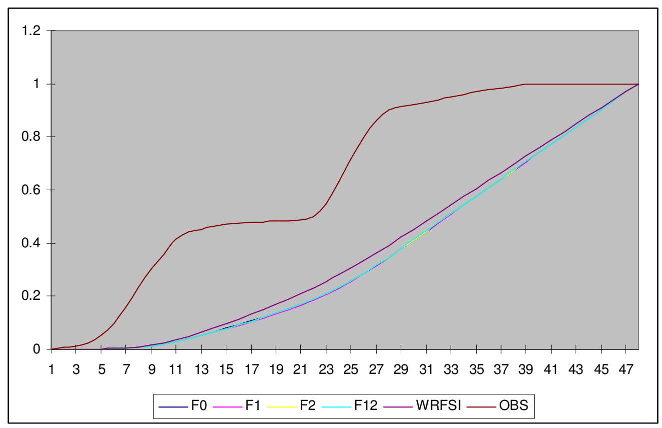


Fig. 7. Normalized accumulated precipitation totals starting at 0GMT May 24th through 0GMT on May 26th for LISWRF integrations and observations.

capture that feature.

## V. GCE MODELING STUDY

In this study a series of 20 day integrations were performed using the GCE model and compared observational cloudy data from ARM. In one integration the observed surface fluxes from the case period were used to drive the GCE model, while in separate integration fluxes from the LIS offline integrations were used for period extending from May 25 through June 14 of 2002. The comparison of LIS generated fluxes to observations is shown in Fig. 8. Overall, LIS shows an excellent agreement with the observations. There is tendency to over predict the sensible heat flux, which can be attributed to dryness and higher temperatures from lack of precipitation in the forcing data mentioned in section III. When LIS surface flux data replace ARM data in the GCE simulations, similar results are obtained except that LIS brings about a better simulation of diurnal cloud variation in the lower troposphere. This suggests that future work with GCE should include more cases to test whether LIS continues to improve the ability of GCE to simulate convection over these long term integrations.

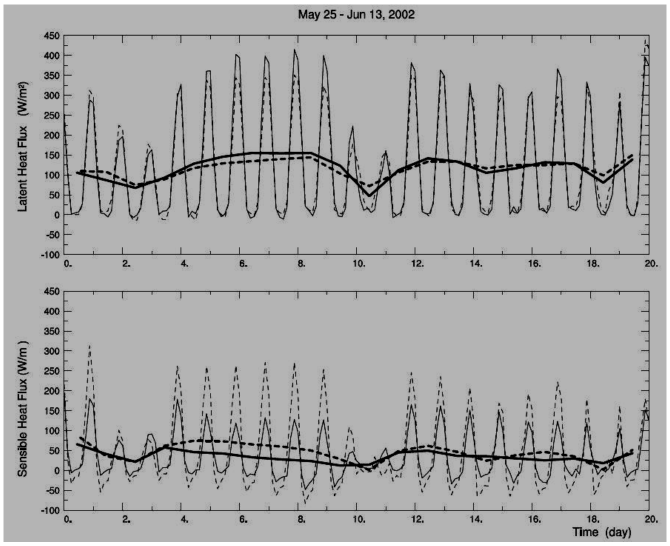


Fig. 8. Time series of surface fluxes for the 2002 case. All data start at 2030 UTC 25 May 2002. Solid and dashed thin lines represent the surface fluxes from the ARM observations and LIS land data assimilation system, respectively. Thick lines represent corresponding daily-averaged values.

## VI. Summary

NASA's Land Information System and the coupling to Weather Research and Forecasting and Goddard Cumulus Ensemble models have undergone further enhancements. The LIS/LISWRF system is now capable of handling multiple projections and now has nesting capabilities. This will aid in the utility of this system for providing optimum land data assimilation for a multitude of models.

In a series of integrations LIS was run offline employing different resolutions of forcing and soil data for input into the LISWRF coupled modeling system. This was done for two case days that represented relatively weak to strong synoptic forcing. These integrations were also contrasted to simulations using the WRFSI system to initialize the land surface initial conditions. These integrations demonstrated the ability of LIS provided ICs to improve the LISWRF models simulated convective activity when compared to the WRFSI initialization. The study also showed that higher resolution forcing and/or soil data does not have a significant impact except in improving the variance captured by the model. The results also suggest that one of the main forcing drivers, precipitation needs to approach scales of the modeled grid.

This work is also demonstrated the ability of LIS fluxes to improve the GCE simulated convective and point to a future need to test this over other case periods. Future work will focus on improving the predictability of LISWRF through employing the nesting capabilities on a continental scale. In addition work is ongoing to produce a high resolution forcing dataset that approaches the scales used in this project.

Acknowledgements

This project is supported by NASA ESTO/AIST NRA-02-OES-04 (PI Peters-Lidard), with additional support from NASA ESTO/CT CAN-00-OES-01 (Co-PI's Houser and Peters-Lidard). This support is gratefully acknowledged.

## References

[1] J.C. Derber, D.F. Parrish and S. J. Lord, 1991: The new global operational analysis system at the National Meteorological Center. *Weather & Forecasting,* 6, 538-547.
[2] M. B.Ek, K. E. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno, and J. D. Tarpley, 2003: Implementation of Noah land-surface model advances in the NCEP operational mesoscale Eta model, *J. Geophys. Res*., 108, No. D22, 8851, doi:10.1029/2002JD003296, 2003.
[3] T.R. Holt, D. Nyogi, F. Chen, K. Manning, M.A. LeMone, and A. Qureshi. Effect of Land-Atmosphere Interactions on the IHOP 24-25 May 2002 Convection Case, *Monthly Weather Review*, 134, 113-133, 2006.
[4] S. Kumar, C. D. Peters-Lidard, Y. Tian, and P. Houser. Early adoption of ESMF by the land information system. In *NASA Earth Science Technology Conference,* College Park, MD, 2003
[5] C. D. Peters-Lidard, S. Kumar, Y. Tian, J.L. Eastman, and P. Houser. Global urban-scale land-atmosphere modeling with the land information system. In *Symposium on Planning, Nowcasting, and Forecasting in the Urban zone, 84th AMS Annual meeting,* Seattle, WA. 2004.
[6] K. E. Mitchell et al., (2004), The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system*, J. Geophys. Res*., 109
[7] NCEP Reanalysis data provided by the NOAA-CIRES ESRL/PSD Climate Diagnostics branch, Boulder, Colorado, USA, from their Web site at http://www.cdc.noaa.gov/
[8] R.A. Pielke Sr., 2005: Land use and climate change. *Science*, 310, 1625-1626.
[9] R.A. Pielke, T.J. Lee, J.H. Copeland, J.L. Eastman, C.L. Ziegler, and C.A. Finley, 1997: Use of USGS-provided data to improve weather and climate simulations. *Ecological Applications*, 7, 3-21.
[10] C.A. Reynolds, T. J. Jackson, and W.J. Rawls. 1999. Estimating Available Water Content by Linking the FAO Soil Map of the World with Global Soil Profile Databases and Pedo-transfer Functions. Proceedings of the AGU 1999 Spring Conference, Boston, MA. May31-June 4, 1999.
[11] M. Rodell, et al. The Global Land Data Assimmilation System. *Bulletin of the American Meteorological Society,* 85(3):381-394, 2004.
[12] B. L. Shaw, R.A. Pielke, and C.L. Ziegler, 1997: A three-dimensional numerical simulation of a Great Plains dryline. *Mon. Wea. Rev*., 125, 1489-1506.
[13] W. C. Skamarock, J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang, and J. G. Powers, 2005: A description of the Advanced Research WRF Version 2. NCAR Tech Notes-468+STR
[14] Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture. State Soil Geographic (STATSGO)
[15] Y. Tian et al., Advanced Computation Technologiesin the GSFC land information system. In *NASA Earth Science Technology Conference,* Palo Alto, CA, 2004.
[16] T. M. Weckworth et al.An overview of the International $H_2O$ Project (IHOP_2002) and some preliminary highlights. *Bull. Amer. Meteor. Soc.,* 85, 253-277.